ABAG

Association of Bay Area Governments

Hotel Claremont • Berkeley, California 94705 • (415) 841-9730

BASIS: THE BAY AREA SPATIAL INFORMATION SYSTEM

THE PROBLEM: DATA FOR PLANNING

Recent years have seen enormous changes in the scope of urban and regional planning. The subject matter of planning, once largely confined to land use and design problems, has spread to areas as diverse as transportation, earthquake hazards, air and water pollution control, criminal justice, housing, and emergency medical services.

This breadth of scope means that the planner needs a massive amount of data about the natural environment (geology, soils, earthquake zones, flooding, etc.) and the patterns of human activity (population, land use, highways, and administrative units). Furthermore, the planner must be able to study the interrelationships among these variables in space and time.

The problem becomes even more difficult if the geographic area of concern is large and diverse. The San Francisco Bay area is comparatively large - 7000 square miles - for a regional planning effort, but still must be studied in some detail because of the heterogeneity of its natural and cultural landscape. The Bay region ranges from large cities to broad areas of open space, from the coastal problems of the ocean to the pressures of suburban growth on prime agricultural land. The result of this size and diversity is clear: more and more data is needed to do meaningful planning.

A further complication is introduced by the political setting; the planning decisions which shape the Bay Area are made by a bewildering number of actors. Cities, counties, the state, the federal government, special districts, regional agencies, corporations, and private land owners all make decisions which influence the environment.

Each of these organizations develops and uses data about the Bay Area to make decisions - and much of that data overlaps data collected and used by other organizations. This duplication, in addition to being costly, results in situations where the decision-making process becomes weighted down with data questions rather than policy issues. Each participant in

a decision must spend substantial time and money resources collecting and analyzing data.

But isn't a lot of this duplication simply different ways of looking at the same basic data ? Since the geographic area of interest is a common denominator, it seems that many of these participants in the planning process could utilize a shared system for collecting and storing that data which they hold in common. Ideally, the participants in any planning decision should have a common data base and be able to address policy implications of the data rather than the data itself. This is a simple concept, but there are formidable barriers - administrative as well as technical - to implementing it.

BASIS

One attempt to apply this concept has been in the development of a system called BASIS - The Bay Area Spatial Information System. BASIS provides a data base containing geographically-referenced data about the San Francisco Bay Area. It was developed by ABAG (the Association of Bay Area Governments, the comprehensive regional planning agency).

BASIS was designed to address the problems outlined above. It is capable of handling a wide variety of environmental and cultural data. It covers the entire nine-county Bay Area. A fundamental concept in its development is that of a shared data base: use of the system is not confined to ABAG projects, but is open to other organizations involved in the planning process.

BASIS is a grid cell system; that is, the geographic area is represented by a rectangular array of square grid cells, and all data is stored and manipulated in terms of these cells. Since the system was designed to be used for local government applications as well as at the regional scale, the grid cells are relatively small in size - 100 meters square, or one hectare. Coverage of the Bay region, including all land area and the Bay itself, requires some two million of these cells. The cell array is based on the standard UTM coordinate system.

BASIS runs on ABAG's Univac (formerly Varian) V-76 minicomputer system. It is a relatively small configuration - 384K bytes of main memory, an 88M byte disk drive, a nine-track tape drive, 300 LPM printer, and six terminals. Two peripheral devices, a digitizer and an electrostatic plotter, were added specifically for handling graphic data. Some of the system characteristics, such as memory size and disk capacity, are very real limitations in handling a large data base of this nature. A further constraint is that the computer is not dedicated to BASIS, but is used for mailing list maintenance, A-95 project notification, a personnel data system, and other applications.

The data base is designed to contain 80 data items for each of the two million hectare cells. TABLE 1 lists the variables now in the data base as well as those in some stage of preparation.

TABLE 1

DATA BASE CONTENTS

BAY AREA SPATIAL INFORMATION SYSTEM

1	**KM ROW NUMBER**	41	(WORKSPACE 1)
2	**KM COL NUMBER**	42	(WORKSPACE 2)
3	TEMPERATURE - MIN	**43**	**COUNTY**
4	TEMPERATURE - MAX	**44**	**1970 CENSUS TRACT**
5	FOG	45	COASTAL/BAY ZONE
6	PREVAILING WIND	**46**	**440 ZONE**
7		47	PUBLIC LAND
8		**48**	**ELEVATION/DEPTH**
9		**49**	**CITY BOUNDARIES**
10		50	SOCIAL PROFILES
11		51	SPECIAL DISTRICTS
12		52	USGS QUAD SHEETS
13		53	LOCAL POLICIES
14		54	ZONING
15		55	LAND USE 1
16		56	LAND USE 2
17		57	LANDSLIDES
18		58	FAULT TRACES
19		**59**	**FAULT ZONES**
20		60	SLOPE
21	**GEOLOGY 1**	61	ASPECT
22	**SOIL TYPES 1**	**62**	**INDUSTRIAL SITES**
23	**USGS FLOOD ZONES**	**63**	**COASTLINE**
24	**PRECIPITATION**	64	HUD FLOOD ZONES
25	**WELL YIELD**	65	WATERSHEDS
26	**GROUND SHAKING**	66	HIGHWAYS
27	**EARTHQUAKE DAMAGE 1**	67	RAIL LINES
28	**EARTHQUAKE DAMAGE 2**	68	NOISE CONTOURS
29	**EARTHQUAKE DAMAGE 3**	69	AIRPORTS
30	**EARTHQUAKE DAMAGE 4**	**70**	**SEAPORTS**
31	**SLOPE STABILITY**	71	MAJOR ACTIVITY SITES
32		72	LANDSCAPE FEATURES
33		73	HISTORIC SITES
34		74	MINERAL RESOURCES
35		75	WILDLIFE HABITATS
36		76	GEOLOGY 2
37		77	SOIL TYPES 2
38		78	UTILITIES
39		79	VEGETATION
40		80	HOSPITALS

This table lists the designated variable spaces within the BASIS data base. Those printed in **BOLDFACE** have been loaded into the data base and edited as of July 1, 1978. The others were in the process of being added at that date or are scheduled for inclusion.

BUILDING THE DATA BASE

Most of the current data base originated in hard copy mapped format. Each data set goes through the following sequence of steps in the data development process.

1) Map Preparation - Base maps are checked for logical consistency, corner points are identified, and data codes are assigned. Disk files containing characteristics of the base maps and data codes are updated if necessary.
2) Digitizing - The mapped data is converted into machine-readable form using an on-line digitizer table and CRT terminal. The mapped lines are treated as chains, or series of points unbroken by vertices. The digitizer operator enters left and right area data codes through the terminal and then follows the map line using a cursor with function buttons. Control information such as map corner points is also entered and stored in disk files along with the chain data.
3) Chain Editing - Working from a plot of the digitized chains, a cartographer checks for correct data codes, topological relationships, and line accuracy. Corrections to the chain file (either data codes or coordinates) are made interactively through a terminal.
4) Chain to Cell Conversion - The digitized chains for a given data set are converted into the BASIS coordinate format (which is basically the standard UTM system, windowed for the Bay region with north-south axis reversed) by mathematical transformation. The transformed chain data is then converted into cell form using a dominant area calculation; i.e., each cell is assigned to the data type which occupies more of the cell than any other data type.
5) Cell Editing - Several types of errors, such as inaccurate following of lines or incorrect assignment of data codes, slip through the chain editing step. At this point in the process, corrections can be made directly to the cell data base, again using an interactive terminal-based program.

Other types of data are available in other than mapped form, and therefore do not go through the data development sequence described above. For example, Landsat imagery and Digital Terrain Tapes are already in machine-readable form, so a different type of conversion process is required. Also, some variables are derived from existing data: a combination of geology and earthquake fault data is used to produce a range of maximum earthquake intensity values.

DATA MANAGEMENT

Considerable attention has been paid to the data management issues in designing the BASIS system. Storing 80 data items for each of two

million grid cells is not a trivial matter on any computer system. Doing it on a minicomputer with a single disk drive and a single tape drive requires a solid data management scheme and clear administrative procedures.

A cell-based system has one very important advantage in managing data: its geographic structure can be viewed as an implicit storage structure. Rows and columns on the ground can be translated into locations on a storage medium such as a disk. Use of a direct access scheme such as this can greatly simplify software and minimize access time. In BASIS, the logical unit of storage is the hectare cell; on disk, however, the physical storage unit is the kilometer cell, consisting of a 10 x 10 array of hectare cells. Each kilometer requires 3600 bytes, representing the 80 data items for the 100 hectares it contains. (To fit 80 items for each of 100 hectares into a 3600 byte record, data codes are packed into four-bit groups using an assembly language routine.)

The mapping of kilometer cells to disk locations is implemented through a disk file which contains a table of location parameters (partition and record number within that partition) for each cell. The location of the data for each hectare cell is implicit within the data space for the kilometer. The ground-to-disk correspondence could, of course, be calculated mathematically from the row and column number without using a table; since the geographical region of concern is not rectangular, however, such a scheme would result in a considerable amount of wasted disk space by reserving room for cells in the ocean or land outside of the nine-county Bay region.

PLOTTING HARDWARE AND SOFTWARE

The ability to produce high quality mapped output is, of course, essential to any system which deals with geographic data. BASIS uses an electrostatic plotter as its primary output device. This type of plotter forms images by placing an electric charge on points that are to be darkened using closely-placed stylii (1/100 inch) and then stepping the paper past an inking mechanism. This method of pattern generation provides very good resolution and appearance in comparison to line printer output. Since it is a non-impact process, a plot can be produced very quickly; this is an important advantage over more conventional pen plotters, particularly for gray-scale shade plots.

Conversion of an image to the corresponding on-off dot pattern for each raster is handled by software. The hardware vendor provides a set of FORTRAN-callable subroutines which performs these functions. Since these routines work by generating vectors and then going through a vector-to-raster conversion step, they are unnecessarily slow for plotting data in grid cell format. Most of these functions have been replaced in the BASIS software package with programs which generate shade or line patterns by turning on specific bits in an output record.

The accompanying example (PLATE 1) illustrates the plotting capabilities necessary for most BASIS output. The typical user will need plots produced at a specific scale (or several different scales) and may also need plots in either a line or shade format. The electrostatic plotter is capable of producing both types of plots, at any reasonable scale. It can also form alphanumeric characters through software or a hardware character generator.

APPLICATIONS: USING THE DATA BASE

BASIS has been used in a number of ways thus far. A summary of major applications will serve to illustrate how such a system is actually used as a tool in solving local and regional planning problems.

The initial use of BASIS was in support of the Environmental Management Plan for the Bay Area. The project objective was very straightforward: to locate areas of land which would provide good sites for disposal of hazardous solid wastes. The emphasis was on environmental data, since factors such as slope stability and flood hazard are clear constraints on the location of this type of facility. A data base of natural factors was built (initially using a cell size of one-quarter square kilometer) and a simple overlay analysis produced maps of relative suitability for disposal sites. This map was then screened manually to arrive at policy recommendations for sites to be studied in more detail.

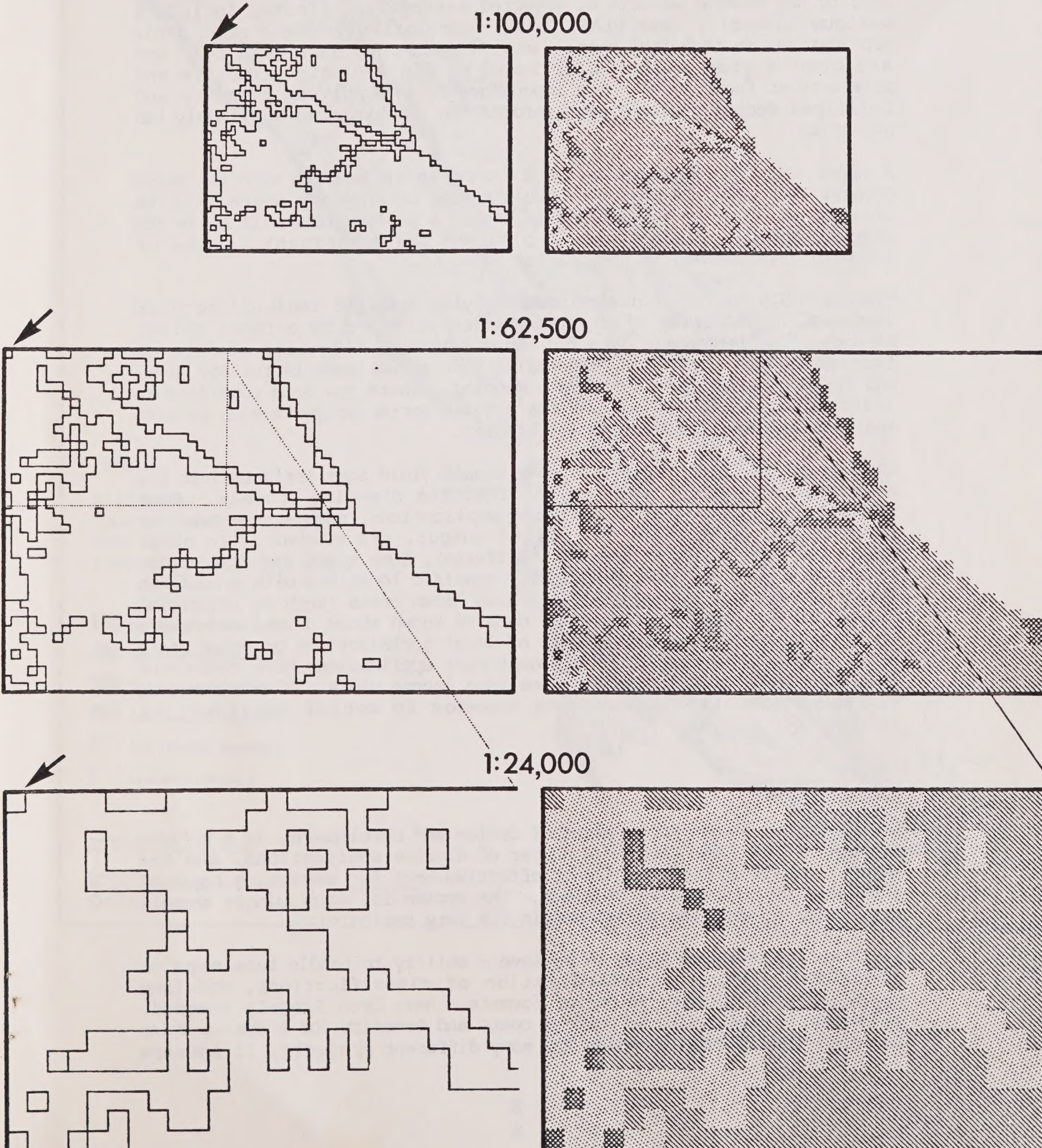
A second application consisted of building a data file about vacant industrial sites in the Bay Area. An overall objective of the project, which was funded by the Governor's Office of Planning and Research, was to look at how environmental concerns could be integrated into the industrial site selection process. Site specific data (lot size and configuration, ownership, availability of utilities, etc.) was obtained through field interviews and then combined with environmental and socioeconomic data using BASIS to produce a detailed listing of data for each site.

In much the same fashion, a file containing information about potential new seaport locations in the Bay was developed for the Metropolitan Transportation Commission. The approach is identical: the universe of sites is predefined, and site specific information is gathered for each location; BASIS is then used as a way to link environmental and socioeconomic data to the sites. Once some spatial attribute (such as coordinates for boundaries or centroid) is generated for each site, it is relatively simple to access the cell data base and selectively add data to the site file.

One of the more impressive uses of the system has been in the area of earthquake hazards. The Bay Area is seismically active, and much research has been done on the theoretical effects of large earthquakes in the region. Little of this work was useful to the planner or concerned citizen, however, because the results could not be readily combined with other types of data to fully assess the human impacts of

PLATE 1

These computer plots are sample maps produced by the BASIS system on an electrostatic plotter. Each map illustrates the same information, either as a line or shade plot, at three different scales. The size of a one hectare cell at each scale is noted by the arrow. The total area covered is 25 square kilometers or about ten square miles.



an earthquake. ABAG, in cooperation with USGS, undertook a program to relate this basic work to the concerns of local govenments. BASIS was used to map several aspects of expected earthquake effects, including maximum intensity (see PLATE 2) and cumulative economic risk. This application, in technical terms, was an easy one; each grid cell was assigned a gray scale value based on its geologic structure and proximity to fault zones. Its importance is in supplying planners and political decision-makers with information to which they previously had no access.

A major current application is a cooperative program with San Mateo County. The county's planners have a number of programs which will be able to utilize an automated data base. A primary role will be in the preparation of a local coastal plan and the subsequent review of proposed developments in the coastal zone.

Planned BASIS applications include supplying data for regional project reviews, an analysis of the population affected by airport noise, mapping of socioeconomic data for human services planning, studies of transportation corridors, a detailed look at wetlands in the Bay Area, and further research on earthquake mapping. There has been considerable interest in using the system for private sector projects such as site analyses and environmental impact studies.

Looking at the uses of BASIS to date should yield some insights into how such systems work and how they fit into the planning process. Some characteristics common to all the applications, such as the need for large amounts of data and for mapped output, are evident. In other ways, the applications are very different. Some start with the entire region as a base and seek to identify specific locations with predefined characteristics, while others start with known areas (such as industrial sites) and aim to collect whatever data is known about them. Most are on a regional scale, but studies of local jurisdictions or other small areas appear to be possible. Although most applications have concerned ABAG projects, other organizations have become users. A wide range of subject areas, from earthquake hazards to social services, is represented.

CONCLUSIONS

BASIS, after almost three years of design and development, is a working system. It has been used in a number of diverse applications, and has proven itself a useful and cost effective tool for addressing regional (and some local) planning problems. The system is, surprisingly enough, very close to what was envisioned in the very beginning.

The benefits which we hoped to achieve - ability to handle base maps of different scales, standardization of classifications, modeling capability, flexibility in output formats - have been largely reached. While it is difficult to compare costs and benefits for a system which has been used by many agencies for many different projects, it appears

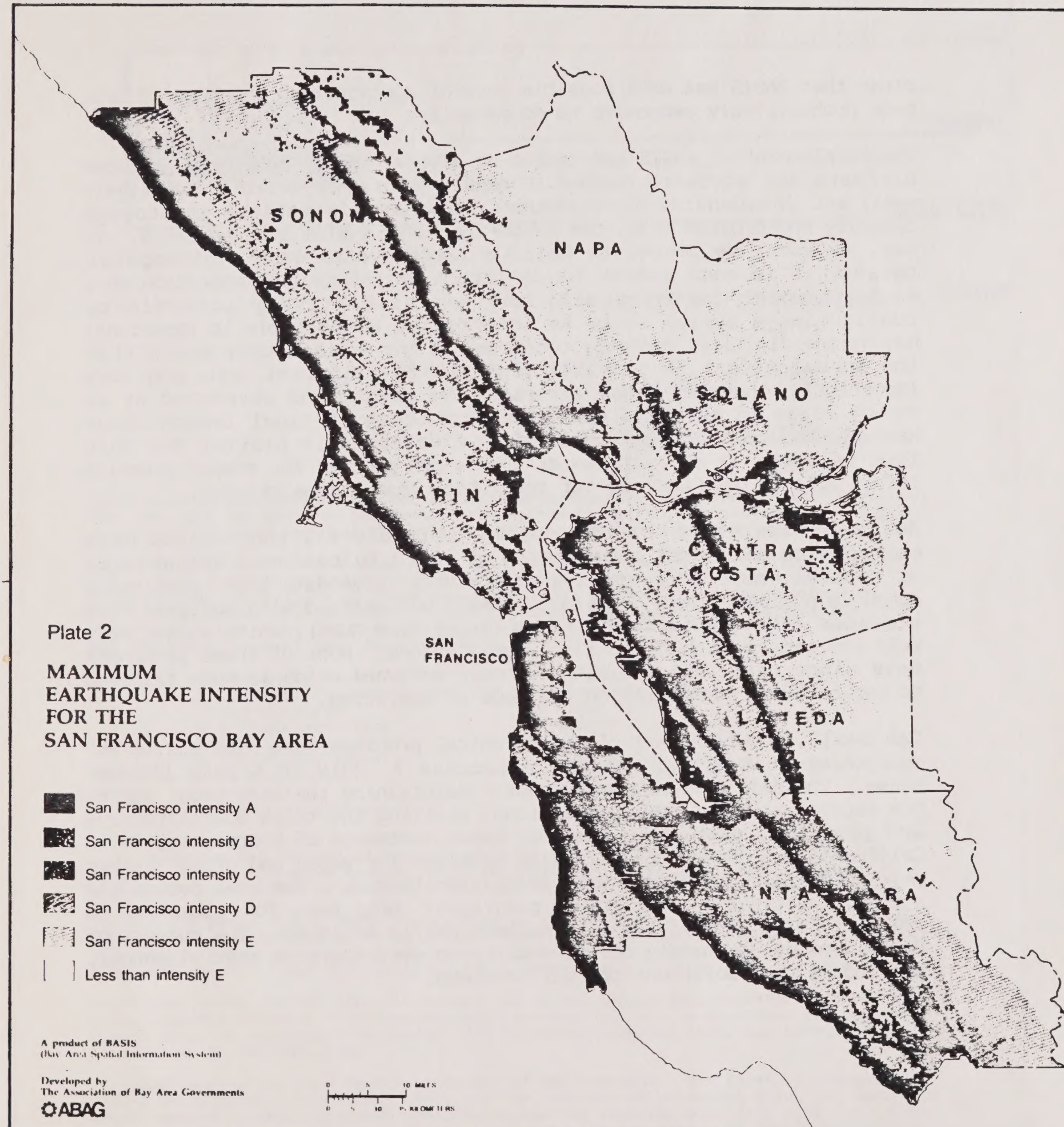


Plate 2

MAXIMUM EARTHQUAKE INTENSITY FOR THE SAN FRANCISCO BAY AREA

clear that BASIS has made possible several applications which would have been prohibitively expensive to do manually.

The development of BASIS has taught us several things about the computer hardware and software needed to handle geographic data. A relatively small and inexpensive minicomputer, despite limitations in storage capacity and program size, can support a large system such as BASIS. It may, in fact, be better to build a large system on a minicomputer because it is much easier to construct an interactive operation on a machine which is physically and administratively accessible. Configuring a system to be as interactive as possible is important: having the digitizer directly connected to the minicomputer means that the manual effort in transferring data is minimized, and, even more importantly, that many input errors can be caught and corrected at an early stage in the data development process. A final lesson about hardware has been the usefulness of an electrostatic plotter for this type of work; it is much faster than a pen plotter for shaded graphics and is able to give much better resolution than a line printer.

Software must be flexible enough to handle all the various data manipulation and output requirements of each application. Reliability is crucial, and, when working with a very large data base, processing speed is obviously very important. BASIS was started with software from a vendor (the GRID and AUTOMAP packages from ESRI), which worked very well and easily handled the first applications. Most of these programs have since been replaced with in-house software which is more tailored to our hardware configuration and mode of operation.

Can BASIS, now that most of the technical problems have been solved and everything is working, be called a success ? Only in a very limited sense. The hard part of the process - maintaining the data base, adding new capabilities for new applications, settling the tough administrative and political problems - has just begun. Passage of Proposition 13 in California has caused major funding problems for BASIS and other public projects which require long-range commitments. The goal behind the system, the creation of a common geographic data base for planning in the San Francisco Bay Area, has been partly achieved. The success of BASIS depends ultimately on its ability to contribute to that planning, in finding real solutions to real problems.

BASIS

BAY AREA SPATIAL INFORMATION SYSTEM

SUMMARY DESCRIPTION

April, 1978

BASIS is a computer-based tool for handling geographically-referenced data. Developed over the past two years by ABAG staff, the system is designed to provide more detailed and timely data for regional and local planning in the San Francisco Bay Area.

BASIS is structured around an array of grid cells, each representing a land area of one hectare (100 meters square in the UTM coordinate system, or about 2.5 acres). It requires over two million of these cells to cover the nine-county Bay region. Each cell on the ground corresponds to a unit of computer storage; the unit contains data codes representing the characteristics of that cell. The system currently contains the following types of data for the region:

- earthquake faults
- ocean and bay coastlines
- precipitation
- geologic materials
- soil associations
- flood prone areas
- slope stability
- well yield
- prime agricultural land
- erosion
- Bay depth contours
- airports
- seaports
- vacant industrial sites
- 1970 Census tracts
- ABAG/MTC 440 zones
- county boundaries

Other types of data (e.g., land use, public land ownership, freeway intersections) are being acquired and encoded as they become available. BASIS is capable of using data based on other coordinate systems (such as longitude/latitude or LANDSAT reference points) by mathematically transforming these reference systems to a common UTM base.

Much of the power of the BASIS system lies in its ability to manipulate the data base. A composite of many data sets can be produced through an overlay process, and can include distance searches or other calculations. Computer-produced maps and/or tabulations are the usual output of the data manipulation process. (See reverse side for sample plots.)

BASIS runs on ABAG's V76 computer system, which can handle data transfer to or from most other computer systems. The computer configuration includes a digitizer for encoding mapped data, an electrostatic plotter for producing computer maps, and terminals for on-line access to the data base.

The BASIS system has been used in a number of applications. The first, a component of ABAG's Environmental Management Plan, was the location of potential sites for hazardous solid waste disposal. A second application was the creation of a data file describing the characteristics of vacant industrial sites in the Bay Area. In ABAG's Earthquake Preparedness Program, BASIS has been used to produce a maximum earthquake intensity map and several cummulative economic risk maps for the entire Bay Area. Current uses include a study of airport noise effects, an analysis of potential seaports, aggregation of population data by special geographic area, and support of San Mateo County's coastal planning program.

For further information on using BASIS, contact Paul Wilson or Don Olmstead at ABAG.

ABAG ASSOCIATION OF BAY AREA GOVERNMENTS
Hotel Claremont · Berkeley, California 94705 · (415) 841-9730

These computer plots are sample maps produced by BASIS and printed on an electrostatic printer-plotter. Each map illustrates the same information, either as an open polygon plot or as a shaded map, at three different scales. The size of a one hectare cell at each scale is noted by the arrow. The area mapped is 25 square kilometers or slightly less than 10 square miles.